Vertical structure and physical processes of the Madden-Julian oscillation: synthesis and summary

Klingaman, N. P., Jiang, X., Xavier, P. K., Petch, J., Waliser, D. and Woolnough, S. J. (2015) Vertical structure and physical processes of the Madden-Julian oscillation: synthesis and summary. Journal of Geophysical Research - Atmospheres, 120 (10). pp. 4671-4689. ISSN 0148-0227 doi: https://doi.org/10.1002/2015JD023196 Available at https://centaur.reading.ac.uk/39952/

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2015JD023196

This article is a companion to *Klingaman et al.* [2015] doi:10.1002/2014JD022374; *Xavier et al.* [2015] doi:10.1002/2014JD022718; and *Jiang et al.* [2015] doi:10.1002/2014JD022375.

Key Points:
- Precipitation-moisture relationships are critical to the simulation of the MJO
- MJO hindcast skill does not equate to fidelity in MJO climate simulations
- Small samples of short hindcasts are difficult to connect to long simulations

Correspondence to:
N. P. Klingaman,
n.p.klingaman@reading.ac.uk

Citation:
Klingaman, N. P., X. Jiang, P. K. Xavier, J. Petch, D. Waliser, and S. J. Woolnough (2015), Vertical structure and physical processes of the Madden-Julian oscillation: Synthesis and summary, *J. Geophys. Res. Atmos.*, *120*, 4671–4689, doi:10.1002/2015JD023196.

Received 30 JAN 2015
Accepted 6 APR 2015
Accepted article online 14 APR 2015
Published online 26 MAY 2015

Vertical structure and physical processes of the Madden-Julian oscillation: Synthesis and summary

Nicholas P. Klingaman[1], Xianan Jiang[2,3], Prince K. Xavier[4], Jon Petch[4], Duane Waliser[2,3], and Steven J. Woolnough[1]

[1] National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK, [2] Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, California, USA, [3] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA, [4] Met Office, Exeter, UK

Abstract The "Vertical structure and physical processes of the Madden-Julian oscillation (MJO)" project comprises three experiments, designed to evaluate comprehensively the heating, moistening, and momentum associated with tropical convection in general circulation models (GCMs). We consider here only those GCMs that performed all experiments. Some models display relatively higher or lower MJO fidelity in both initialized hindcasts and climate simulations, while others show considerable variations in fidelity between experiments. Fidelity in hindcasts and climate simulations are not meaningfully correlated. The analysis of each experiment led to the development of process-oriented diagnostics, some of which distinguished between GCMs with higher or lower fidelity in that experiment. We select the most discriminating diagnostics and apply them to data from all experiments, where possible, to determine if correlations with MJO fidelity hold across scales and GCM states. While normalized gross moist stability had a small but statistically significant correlation with MJO fidelity in climate simulations, we find no link with fidelity in medium-range hindcasts. Similarly, there is no association between time step to time step rainfall variability, identified from short hindcasts and fidelity in medium-range hindcasts or climate simulations. Two metrics that relate precipitation to free-tropospheric moisture—the relative humidity for extreme daily precipitation and variations in the height and amplitude of moistening with rain rate—successfully distinguish between higher-fidelity and lower fidelity GCMs in hindcasts and climate simulations. To improve the MJO, developers should focus on relationships between convection and both total moisture and its rate of change. We conclude by offering recommendations for further experiments.

1. Introduction

Many contemporary general circulation models (GCMs) used for numerical weather prediction and climate simulations fail to capture the defining characteristics of the Madden-Julian oscillation (MJO) [*Madden and Julian*, 1971, 1972]: its 30–70 day period, the zonal wave number 1 structure of its circulation in the deep tropics, and the approximately 5 m s^{-1} eastward propagation speed of convective anomalies from the Indian Ocean through the Maritime Continent to the West Pacific [e.g., *Lin et al.*, 2008; *Hung et al.*, 2013]. *Zhang* [2005] reviews the MJO and its global teleconnections. The "Vertical structure and physical processes of the Madden-Julian oscillation" global-model evaluation project seeks to explore the underlying causes of these deficiencies, which are commonly thought to lie in GCM subgrid physical parameterizations and the interactions between those parameterizations and the resolved GCM dynamics. The project is organized and supported by the WCRP-WWRP/THORPEX-YOTC MJO Task Force and the GASS panel of GEWEX (The acronyms refer to the World Climate Research Programme (WCRP), the World Weather Research Programme (WWRP), The Observing System Research and Predictability Experiment (THORPEX), the Years of Tropical Convection (YOTC), the Global Atmospheric Systems Studies (GASS) panel, and the Global Energy and Water Exchanges Project (GEWEX). At the time of this study, the MJO Task Force was under the collective auspices of WCRP, WWRP, THORPEX, and YOTC. It is currently under the auspices of the Working Group on Numerical Experimentation (WGNE).). The overall aim of the project is to characterize, compare, and evaluate the heating and moistening processes associated with the MJO in GCMs, with a particular focus on the vertical structures of those processes [*Petch et al.*, 2011]. Such detailed evaluations of simulated processes are possible only because of the substantial quantity of satellite-derived observations and high-resolution analyses collected during YOTC [*Moncrieff et al.*, 2012; *Waliser et al.*, 2012].

The project comprises three experiments, designed to take advantage of links between GCM biases in initialized hindcasts and those in free-running, multidecadal climate simulations. These links have been demonstrated in previous model intercomparison projects, such as the Transpose Atmospheric Model Intercomparison Project II (Transpose-AMIP2) [*Williams et al.*, 2013] as well as separate efforts by individual groups [e.g., *Boyle et al.*, 2008]. The three experiments are (a) multidecadal, present-day control climate simulations ("20 year climate simulations"); (b) very short initialized hindcasts of two strong, well-observed MJO events in boreal winter 2009–2010 ("2 day hindcasts"); and (c) medium-range hindcasts of the same MJO events with a greater range of initial dates ("20-day hindcasts"). Each component collected frequent output, ranging from time step data for (b) to 6 h data for (a), of tropospheric vertical profiles of prognostic variables and tendencies from individual subgrid-scale parameterizations, to create a rich and comprehensive data set spanning spatial and temporal scales. We describe and analyze these experiments in separate manuscripts: *Jiang et al.* [2015], *Xavier et al.* [2015] and *Klingaman et al.* [2015] for (a), (b), and (c) above, respectively. Some details on the experiments and their objectives are given in section 2.1, but the reader is encouraged to refer to the above studies.

Since the project emphasized links between GCM behavior at short and long temporal scales, some additional analysis is warranted to extend the findings from each experiment. In particular, *Jiang et al.* [2015] and *Klingaman et al.* [2015] each identified "process-oriented" diagnostics that the authors found were able to distinguish between GCMs that produced relatively more or less accurate representations of the MJO; hereafter, we refer to this as "MJO fidelity." GCMs with high MJO fidelity in 20 year climate simulations exhibited strong sensitivity of precipitation to free-tropospheric relative humidity, as well as a tendency for a positive feedback between anomalies in convection and moist static energy via convection-induced vertical circulations. GCMs with high MJO fidelity in the 20 day hindcasts showed a smooth increase in the height of moistening (i.e., positive tendencies of moisture) with precipitation, with low-to middle-tropospheric moistening at moderate rain rates particularly important. In the 2 day hindcasts, *Xavier et al.* [2015] demonstrated that GCM biases in temperature and moisture developed quickly, driven by large model-to-model variability in the amount and type of clouds and their interactions with radiation. Several GCMs also showed substantial time step to time step variability in convection and consequently poor convection-dynamics coupling at the time step level. *Xavier et al.* [2015] alone, however, it was not possible to determine whether these GCM behaviors extended to the other temporal scales captured in the project.

Here we apply the most discriminating and revealing diagnostics developed in each component of the project to the data collected in the other experiments, where possible. Of the 32 GCM contributions received—counting multiple configurations of the same GCM separately—27 GCMs performed climate simulations, 14 performed 20 day hindcasts, and 12 performed 2 day hindcasts. For consistency, we use only the set of nine GCMs that contributed results to all three experiments (section 2.2). This allows us to achieve our objective of examining GCM behavior across temporal scales, which requires all three experiments, as well as to display and discuss results from all GCMs considered.

We have tried to strike a balance between providing the detail needed to understand our diagnostics and repeating information from the other three manuscripts. We give limited information on the data used and calculations performed, such that it should be possible to understand our results using only the information given here, but the reader is strongly encouraged to consider all four manuscripts as a set. We briefly describe the data used in this study (section 2), apply diagnostics developed for one experiment to data from the others (section 3), and discuss (section 4) and summarize (section 5) the results from the project.

2. Models and Methods

2.1. Experiments

Modeling centers were asked to perform the three experiments with either atmosphere-ocean coupled or atmosphere-only GCMs. The 20 year climate simulations are designed to measure MJO fidelity when the GCM is close to its mean climate; the 2 day hindcasts aim to evaluate parameterization behavior when all GCMs are constrained by a common initial analysis; the 20 day hindcasts bridge the gap between the other experiments by linking hindcast MJO fidelity to biases in physical processes as the GCMs drift toward their preferred states. Hindcasts were initialized daily during two MJO events in YOTC from European Centre for Medium-range Weather Forecasts YOTC (ECMWF-YOTC) 00Z analyses: the 20 day (2 day) hindcasts were initialized on 10 October 2009 to 25 November 2009 (20 October 2009 to 10 November 2009) and 10 December 2009 to

Table 1. For Each GCM: The Full Name, Version, Abbreviation Used in the Text, Code Used in the Figures, Contributing Center, Native Horizontal Resolution (Longitude × Latitude or Wave Number Truncation (T) With Equivalent in Degrees) and Number of Vertical Points (L), Time Step, and a Reference With Details

Model Name	Ver.	Abbrev.	Code	Center	Resolution	Time Step	Reference
Community Atmospheric Model[a]	5	CAM5	C5	NCAR[b]	1.25° × 0.9°, L30	30 min	*Neale et al.* [2012]
CAM5 with convective microphysics[a]	5.1	CAM5-ZM	CZ	UCSD[c], LLNL[d]	1.25° × 0.9°, L30	30 min	*Song et al.* [2012]
Canadian Coupled Model	4	CanCM4	CC	CCCma[e]	T63 (1.9°), L35	60 min	*Merryfield et al.* [2013]
CNRM[f] Atmospheric Model	5.2	CNRM-AM	CN	CNRM[f]	T127 (1.4°), L31	30 min	*Voldoire et al.* [2013]
European Community Model	3	ECEarth3	E3	SMHI[g]	T255 (0.7°), L91	45 min	*Hazeleger et al.* [2012]
Goddard Earth Observing System GCM	5	GEOS5	NA	NASA[h]	0.625° × 0.5°, L72	20 min	*Rienecker et al.* [2008]
Goddard Institute for Space Studies GCM	E2′	GISS-E2	GI	NASA[h]	2° × 2.5°, L40	30 min	*Schmidt et al.* [2014]
Met Office Unified Model	GA3	MetUM-GA3	MO	MOHC[i]	0.83° × 0.56°, L70	12 min	*Walters et al.* [2011]
Model for Interdisciplinary Research on Climate	5	MIROC5	MI	AORI, NIES, JAMSTEC[j]	T85 (1.4°), L40	12 min	*Watanabe et al.* [2010]
MRI[k] Atmospheric GCM	3	MRI-AGCM3	MR	MRI[k]	TL159 (1.1°), L48	30 min	*Yukimoto et al.* [2012]

[a]CAM5 2 day hindcasts from CAM5 are combined with CAM5-ZM 20 day hindcasts and 20 year climate simulations.
[b]National Center for Atmospheric Research.
[c]University of California, San Diego.
[d]Lawrence Livermore National Laboratory.
[e]Canadian Centre for Climate Modelling and Analysis.
[f]Centre National de Recherches Météorologiques.
[g]Swedish Meteorological and Hydrological Institute.
[h]National Aeronautics and Space Administration.
[i]Met Office Hadley Centre.
[j]Atmosphere-Ocean Research Institute (AORI), National Institute for Environmental Studies (NIES), and Japan Agency for Marine Earth Technologies (JAMSTEC).
[k]Meteorological Research Institute.

25 January 2010 (20 December 2009 to 10 January 2010). Alongside standard prognostic and surface fields, all experiments obtained tendencies from individual GCM parameterizations and the GCM dynamics for temperature, specific humidity, and zonal and meridional winds. The 20 year climate simulations and 20 day hindcasts collected output either globally (for prognostic and surface fields) or across 50°S – 50°N (for tendencies) on a 2.5° × 2.5° grid and 24 pressure levels; 6 h (3 h) data were provided for the 20 year climate simulations (20 day hindcasts). For 2 day hindcasts, centers provided time step data on the GCM native horizontal and vertical grid over 60° – 160°E, 10°S – 10°N.

2.2. Models

We analyze data from the nine GCMs that contributed to all three experiments (Table 1). In the text, we refer to GCMs by the "Abbreviation" in Table 1; in figures, we label GCMs with the two-letter "Code." The remainder of this section lists exceptions or caveats to the data and our analysis, to which the reader is encouraged to refer in conjunction with the analysis in section 3.

While each center provided a reference for their GCM (Table 1), certain deviations from these or from the experiment design were made that are relevant to our analysis. CanCM4 is the only coupled GCM in this set of nine models. To maintain atmosphere-ocean coupled balance, the CanCM4 2 day and 20 day hindcasts were initialized from analyses generated by the operational coupled assimilation system from the Canadian Centre for Climate Modelling and Analysis, rather than ECMWF-YOTC analyses. *Klingaman et al.* [2015] and *Xavier et al.* [2015] discuss the implications for this discrepancy on CanCM4 hindcast fidelity. CAM5 and CAM5-ZM used finite-volume dynamical cores. The GISS-E2′ ("E2 prime") convection scheme was modified from version E2 to improve tropical intraseasonal variability [*Del Genio et al.*, 2012; *Kim et al.*, 2012]. For the 2 day and 20 day hindcasts, the choice of SST boundary condition was left to the centers. Among the atmosphere-only models, MRI-AGCM3 and CNRM-AM persisted the initial SST; ECEarth3, MetUM-GA3, and GISS-E2′ persisted the initial SST anomaly with respect to a time-varying climatology; MIROC5, CAM5-ZM, and GEOS5 used time-varying observed SSTs. Previous studies have found that using high frequency, observed SSTs increase MJO prediction skills [e.g., *Kim et al.*, 2008; *de Boisséson et al.*, 2012], but this increase is artificial because it provides the model with information not available at the start of the hindcast. In the set of GCMs used here, *Klingaman et al.* [2015] found no correlation between the choice of SST boundary condition and hindcast MJO fidelity.

For the analysis of moistening tendencies by rain rate (section 3.5), we combine results from two configurations of CAM5: the standard CAM5 and CAM5-ZM, which add the *Song and Zhang* [2011] convective microphysics to the standard *Morrison and Gettelman* [2008] stratiform scheme. We use CAM5-ZM moistening tendencies from the 20 year climate simulations and 20 day hindcasts, as these data were missing or incomplete from CAM5. CAM5-ZM did not perform 2 day hindcasts, so we use CAM5 tendencies instead. *Klingaman et al.* [2015] and *Jiang et al.* [2015] demonstrated that CAM5 and CAM5-ZM performed similarly in 20 day hindcasts and 20 year climate simulations, respectively, so this substitution should not affect our results. Further, the data required for the relative humidity difference (section 3.2) and normalized gross moist stability (section 3.3) metrics were not archived from the MetUM-GA3 20 year climate simulation. Although data from the Superparameterized CAM (SPCAM3) were submitted to all experiments, we exclude SPCAM3 because a different version of the model was used for the 20 year climate simulations than for the 20 day and 2 day hindcasts.

2.3. Data

In section 3.5, we compare GCM relationships between moistening tendencies and rain rates to the same relationship in ECMWF-YOTC 24 h forecasts for the 20 day hindcast period: 10 October 2009 through 15 February 2010. These short forecasts use the ECMWF Integrated Forecast System (IFS; cycle 35r3) at 16 km horizontal resolution, initialized from the ECMWF-YOTC analyses (also from IFS 35r3) used for the 2 day and 20 day hindcasts. While the IFS parameterizations undoubtedly influence the structure and amplitude of the moistening tendencies, at these short lead times the model should be reasonably well constrained at larger scales by the analyses. In the absence of direct observations of moistening, we consider ECMWF-YOTC the closest available approximation to reality. ECMWF-YOTC net moistening was computed by summing the tendencies from the individual IFS physics schemes together with the tendency from the dynamics; this avoids any influence from analysis increments. The 3 h ECMWF-YOTC data were interpolated to $2.5° \times 2.5°$ horizontal resolution to agree with data from the 20 day hindcasts and 20 year climate simulations.

3. Results

3.1. MJO Fidelity in Climate and Hindcast Simulations

Jiang et al. [2015] and *Klingaman et al.* [2015] showed qualitatively that there is little correspondence between MJO fidelity in the 20 year climate simulations and 20 year hindcasts. *Jiang et al.* [2015] measured MJO fidelity by pattern correlations of rainfall Hövmoller diagrams between each GCM and observations. The Hövmoller diagrams were constructed from regressions of latitude-averaged (5°S–5°N), 20–100 day band-pass-filtered precipitation on two base regions: the Indian Ocean (75°–85°E, 5°S–5°N) and the West Pacific (130°–150°E, 5°S–5°N). The MJO fidelity score is the mean of the two pattern correlation coefficients. In the 20 day hindcasts, *Klingaman et al.* [2015] assessed MJO fidelity as the first lead time at which the bivariate correlation of the simulated and observed *Wheeler and Hendon* [2004] Real-time Multivariate MJO (RMM) indices was less than a subjectively chosen critical value of 0.7. The RMM skill measure agreed reasonably well, but not perfectly, with pattern correlations of rainfall Hövmoller diagrams from the models, constructed at fixed hindcast lead times from unfiltered daily means, with similarly constructed Hövmollers from TRMM. These pattern correlations approximated the method of *Jiang et al.* [2015] to the maximum extent possible, given the limited length of the hindcasts.

The nine GCMs examined here span the range of fidelity found for all GCMs in the 20 day hindcasts and 20 year climate simulations: GISS-E2 and MRI-AGCM3 (CAM5-ZM and GEOS5) were among the highest-fidelity models in the 20 year climate simulations (20 day hindcasts), while CanCM4 and MetUM-GA3 (CanCM4 and MIROC5) were among the lowest-fidelity models in the 20 year climate simulations (20 day hindcasts); ECEarth3 and CNRM-AM performed moderately well in both components (Table 2). For these nine GCMs, there is no useful relationship between these fidelity measures (Figure 1a). Although the two measures are correlated ($r = 0.55$, $p \sim 0.15$), if CanCM4 is removed the correlation is weakened substantially ($r = 0.32$, $p > 0.20$), and the regression line becomes nearly vertical. GISS-E2 and MRI-AGCM3 increase in relative performance between the 20 day hindcasts and 20 year climate simulations, while CAM5-ZM, GEOS5, and MetUM-GA3 show declines in fidelity; MIROC5, CanCM4, CNRM-AM, and ECEarth3 perform similarly in the two experiments, relative to the full set of GCMs in each component. MJO performance in medium-range hindcasts does not necessarily translate to performance in multidecadal climate simulations, or vice versa, at least not for these GCMs and fidelity measures.

Table 2. For Each Model, the Pattern Correlations of Hövmoller Diagrams of Daily Mean Rainfall, Averaged 10°S–10°N, Are Computed Against Similarly Constructed Diagrams From TRMM[a]

	Day 2 (60°–160°E)		Day 2 (60°–105°E)		Day 10 (to 160°E)		Day 10 (to 105°E)		Classifications	
Model	Corr	Rank	Corr	Rank	Corr	Rank	Corr	Rank	20 day	20 year
CNRM-AM	0.637	5/9	0.655	5/9	0.008	9/9	-0.193	9/9	Moderate	Moderate
MetUM-GA3	0.615	7/9	0.600	8/9	0.256	5/9	0.107	6/9	Moderate	Lower
ECEarth3	0.656	3/9	0.671	4/9	0.176	7/9	0.093	7/9	Moderate	Moderate
MRI-AGCM3	0.674	1/9	0.684	2/9	0.280	3/9	0.168	4/9	Moderate	Higher
CAM5-ZM	0.653	4/9	0.705	1/9	0.266	4/9	0.173	3/9	Higher	Moderate
MIROC5	0.633	6/9	0.633	6/9	0.187	6/9	0.127	5/9	Lower	Lower
GEOS5	0.659	2/9	0.681	3/9	0.396	1/9	0.190	2/9	Higher	Moderate
CanCM4	0.496	9/9	0.492	9/9	0.104	8/9	0.064	8/9	Lower	Lower
GISS-E2	0.593	8/9	0.601	7/9	0.395	2/9	0.349	1/9	Moderate	Higher

[a]Correlations are computed from 20 day hindcast data, using only start dates from the 2 day hindcast experiment, at 2 day and 10 day lead times. The 10 day lead time is the same diagnostic used in *Klingaman et al.* [2015], except for only the 2 day hindcast start dates rather than all 20 day hindcast start dates. Pattern correlations are computed over both 60–160°E and 60–105°E. For each set of pattern correlations, the rank of each model is shown for clarity. The classifications of each model in the 20 day hindcasts and 20 year climate simulations are shown in the far-right columns, using the color-coded system described in the text, for comparison with the ranks.

In Figure 1 and throughout, we color models' codes by their MJO fidelity relative to all GCMs that submitted data for that experiment, not only the nine GCMs shown here. For the 20 year climate simulations, we follow the quartile classifications in *Jiang et al.* [2015]: the "top 25%" (upper quartile) of models are colored red; the middle 50% are black; the "bottom 25%" (lowest quartile) of models are blue. For the 20 day hindcasts, we follow the tercile classifications in *Klingaman et al.* [2015], with the same color scheme as in *Jiang et al.* [2015]. We note that the classification boundaries do not align between the experiments, but that aligning them would not change our conclusions.

Xavier et al. [2015] did not assess MJO fidelity in the 2 day hindcasts, not only because of the limited length and sample size of the hindcasts but also because the focus of that study was on the short-range, time step behavior of model physical and dynamical processes, rather than on MJO skill. Here we attempt to connect MJO fidelity in the 2 day hindcasts to the 20 day hindcasts and 20 year climate simulations by calculating pattern correlations of longitude-time rainfall Hövmoller diagrams between each model at a 2 day lead time (hours 25–48) and TRMM. We construct the Hövmollers as in *Klingaman et al.* [2015] and compute the pattern correlation over two longitude bands: 60°–160°E, the full domain of the 2 day hindcasts; and 60°–105°E, which is approximately the domain of the active MJO phase in the 2 day hindcast cases [see Figure 1 in *Xavier et al.*, 2015]. We use rainfall data from the 20 day hindcasts for convenience; the 2 day hindcasts are identical to the first two days of the 20 day hindcasts due to the use of the same models and initial conditions. Pattern correlations are computed over all 2 day hindcast start dates for each case, then averaged between the two cases. We also compute the correlations at a 10 day lead time, again using data from the 20 day hindcasts.

Table 2 lists the pattern correlations, ranks the models, and includes for comparison the classifications of the models in the 20 day hindcasts and 20 year climate simulations. With the exception of CanCM4, all models produce highly similar pattern correlations at 2 day lead times for both longitude bands. By this measure of MJO fidelity, there is little difference among the models at such short lead times. The pattern correlations at 2 day leads show little correspondence to the pattern correlations at 10 day leads: GISS-E2 has a relatively low correlation at 2 day leads but a relatively higher correlation at 10 day leads; ECEarth3 displays the opposite behavior. Further, the pattern correlations at 2 day leads are poor predictors of MJO fidelity in the 20 day hindcasts and 20 year climate simulations, as shown by the classifications in the right-hand column. Figure 1b confirms that there is no statistically significant correlation ($r = 0.27, p > 0.20$) between the pattern correlations at 2 day and 10 day leads. Similarly, Figure 1c demonstrates that there is a weak and insignificant correlation ($r = 0.50, p \sim 0.15$) between the pattern correlations at 2 day lead time and MJO fidelity in the 20 year climate simulations. Models that perform similarly well at 2 day lead times, such as CNRM-AM, MIROC5, and CAM5-ZM, show highly disparate fidelity in 20 day hindcasts and 20 year climate simulations. We discuss hypotheses

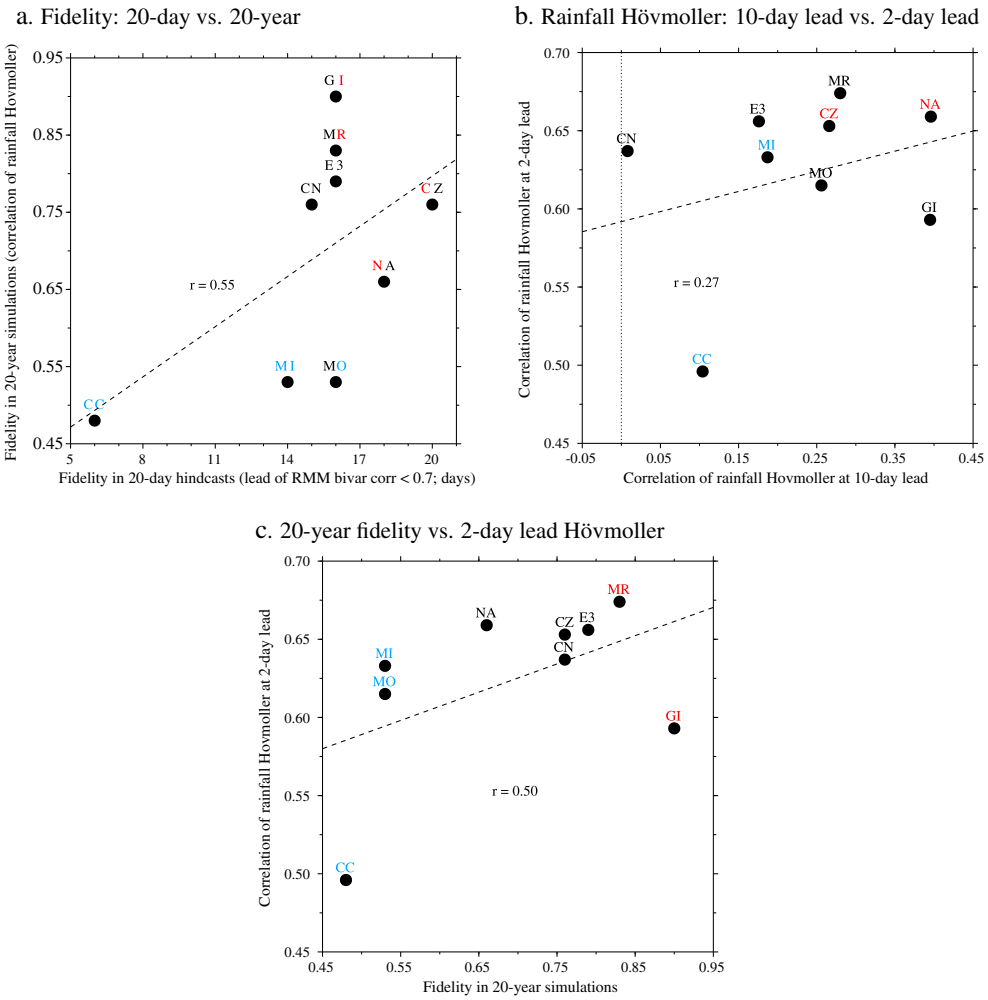

Figure 1. For each model, (a) MJO fidelity in 20 day hindcasts against fidelity in 20 year climate simulations; (b) pattern correlations of Hövmoller diagrams of daily mean rainfall (averaged 10°S–10°N and computed over 60°–160°E) between TRMM and the model, at 2 day and 10 day lead times; (c) fidelity in 20 year climate simulations against the 2 day pattern correlations from Figure 1b. In Figure 1a, the color of the first letter of each code gives relative fidelity among all thirteen 20 day hindcast models (blue: lower tercile; black: middle tercile; red: upper tercile); the color of the second letter gives relative fidelity among all twenty-seven 20-year climate simulation models (blue: lower 25%; black: middle 50%; red: upper 25%). In Figure 1b, the codes show 20 day hindcast fidelity; in Figure 1c, they show fidelity in 20 year climate simulations. The least-squares regression lines and correlation coefficients are also shown. Without CanCM4 (CC) the correlation in Figure 1a is 0.32.

for the disconnects in estimated MJO fidelity among the three experiments in section 4. Because of the similarity in correlation values and the very limited sample of start dates available, we do not separate the 2 day hindcasts based on MJO fidelity; in the figures, all model codes are colored black for relatively "moderate" fidelity.

3.2. Relationship Between Precipitation and Relative Humidity

Using eight GCMs, *Maloney et al.* [2014] demonstrated a positive correlation between fidelity in East Pacific subseasonal variability and the difference in average lower tropospheric to middle (850–500 hPa) tropospheric relative humidity (RH) for the heaviest and lightest daily rain rates. This metric measures the sensitivity of simulated precipitation to moisture; various alternate forms have been proposed recently by the MJO community [e.g., *Thayer-Calder and Randall*, 2009; *Xavier*, 2012; *Kim et al.*, 2014]. For each of the 27 GCMs for which 20 year climate simulations were submitted, *Jiang et al.* [2015] computed the difference in mean mass-weighted 850–500 hPa RH for the heaviest 5% and lightest 10% daily rain rates across a Warm Pool domain (60°E–180° and 15°S–15°N). The authors found a moderately strong correlation ($r = 0.45$, $p \sim 0.02$) between this metric and MJO fidelity.

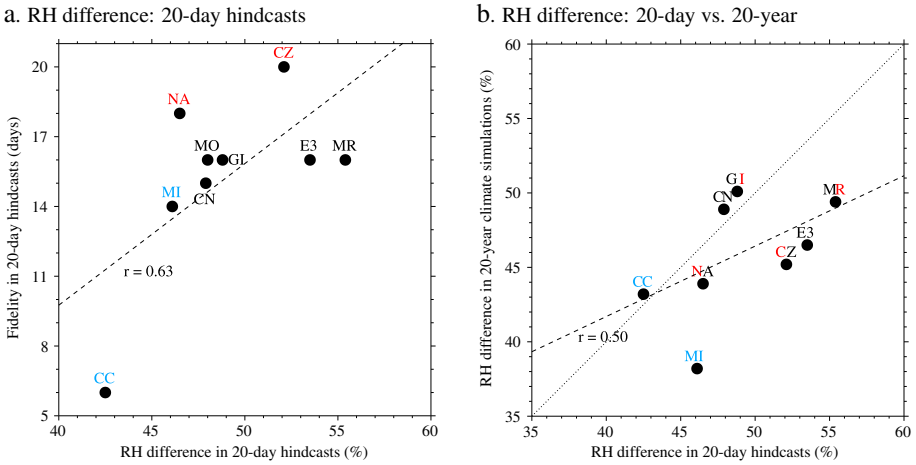

Figure 2. For each model, the difference in mass-weighted 850–500 hPa averaged relative humidity between the heaviest 5% and lightest 10% of daily rainfall events in the 20 day hindcasts (using days 3–20 only), against (a) MJO fidelity in the 20 day hindcasts and (b) the same difference in relative humidity from the corresponding 20 year climate simulations. The least-squares regression lines and correlation coefficients are also shown; Figure 2b also includes a one-to-one line. In Figure 2a, model codes are colored by fidelity in 20 day hindcasts only, relative to all thirteen 20 day hindcast models. In Figure 2b, model codes are colored as in Figure 1. Required data for this diagnostic were not archived from the MetUM-GA3 20 year climate simulation, so that model is excluded from Figure 2b.

We compute this metric (hereafter "the RH difference metric") for the 20 day hindcasts from the nine GCMs considered here, combining days 3–20 for all start dates. We exclude the first 48 h of each hindcast due to large trends in RH as the GCM adjusts its column moisture away from the ECMWF-YOTC analysis. For this reason, we do not compute the RH difference metric for the 2 day hindcasts. We also computed the RH difference metric for days 11–20 only, but found only very small differences (±2% RH at most) in model scores, suggesting little variation in the metric with lead time after the first two days. There is a moderately strong relationship ($r = 0.63$, $p \sim 0.10$) between the RH difference metric and MJO fidelity in the 20 day hindcasts that is statistically significant at the 10% level (Figure 2a). However, there are five models with similar values of the RH difference metric—GEOS5, MetUM-GA3, GISS-E2, CNRM-AM, and MIROC5—but with substantial differences in MJO fidelity in the 20 day hindcasts. While there is a positive overall relationship between the RH difference metric and MJO fidelity in the 20 day hindcasts and 20 year climate simulations, the metric does not discriminate perfectly between higher-fidelity and lower fidelity models.

Klingaman et al. [2015] found no statistically significant relationship in the 20 day hindcasts between MJO fidelity and the pattern correlation of specific humidity anomalies as a function of precipitation rate between each model and ECMWF-YOTC 24 h forecast data, which is seemingly at odds with our above result for the RH difference metric. However, *Klingaman et al.* [2015] considered the full vertical profile of specific humidity anomalies, as well as all precipitation rates, whereas the RH difference metric focuses on extreme precipitation rates at both ends of the spectrum and only 850–500 hPa. We hypothesize that it is the focus on precipitation extremes and the lower troposphere that leads to a stronger connection to MJO fidelity for the RH difference metric than for the specific humidity metric in *Klingaman et al.* [2015].

Next, we compare the RH difference metrics in the 20 day hindcasts and 20 year climate simulations, to assess whether, for a single GCM, variations in the metric can explain variations in MJO fidelity (Figure 2b). CanCM4 and MIROC5 show relatively low fidelity in both experiments and produce relatively low values of the RH metric. From the 20 day hindcasts to the 20 year climate simulations, CAM5-ZM displays moderate reductions in both MJO fidelity (Figure 1) and the RH difference metric. Yet GEOS5 also loses fidelity in the 20 year climate simulations relative to the 20 day hindcasts but has only a slight decrease in the RH difference metric. Of the two models that increase in fidelity between the 20 day hindcasts and 20 year climate simulations, GISS-E2 displays an increase in RH difference while MRI-AGCM3 shows a decline, although MRI-AGCM3 still has one of the highest RH difference metric values among all 20 year climate simulations. The RH difference metric scores in the two experiments are only modestly correlated ($r = 0.50$, $p \sim 0.20$) and variations in these scores are only sometimes able to account for variations in MJO fidelity. All GCMs produce RH difference metrics in the 20 year climate simulations that are either less than or similar to their RH difference metrics in the 20 day

hindcasts. This suggests that when the GCM mean state is closer to observations, either the GCM precipitation is more sensitive to RH or the GCM dynamic range of RH is greater.

3.3. Normalized Gross Moist Stability

Gross moist stability (GMS) essentially describes how efficiently convection and divergent flows remove moisture from an atmospheric column, relative to the import of moisture by advection [*Neelin and Held*, 1987; *Raymond et al.*, 2009]. Several studies have hypothesized that a strong MJO is associated with a negative GMS—a positive feedback in which the active (suppressed) MJO induces a circulation that further moistens (dries) the column—in GCMs and in reality [e.g., *Hannah and Maloney*, 2011; *Benedict et al.*, 2014; *Pritchard*, 2014]. *Jiang et al.* [2015] computed the winter (November–April) normalized GMS (NGMS) over ocean points in an Indo-Pacific domain (60°–150°E, 15°S–15°N) following the method in *Benedict et al.* [2014], to which the reader should refer for further details. For all 27 GCMs that performed 20 year climate simulations, there were small but statistically significant negative correlations between MJO fidelity and both vertical NGMS ($r = -0.36$, $p \sim 0.10$) and total NGMS ($r = -0.46$, $p \sim 0.02$).

We calculate NGMS from the 20 day hindcasts as in *Jiang et al.* [2015], using all start dates but only days 3–20, as for the RH difference metric. When we computed NGMS as a function of lead time, by concatenating all hindcast cases at a fixed lead time, most GCMs showed large-amplitude NGMS values—either positive or negative—in the first two days. This suggests strong effects of spin-up on NGMS and prevented us from computing NGMS for the 2 day hindcasts. Since the *Benedict et al.* [2014] procedure requires a 17 day smoothing, we obtain one NGMS value from each 20 day hindcast, using days 3–20 and daily means. We then average NGMS across all 94 hindcasts. We stress that this is an extremely small sample of data from which to compute NGMS, which exhibits large variability from one grid point and day to the next. NGMS calculations are usually performed on at least a decade of model data or observations, which span a range of synoptic conditions. We have only 94 days of data, after temporal smoothing, and most of the 20 day hindcast start dates include an active MJO. The results presented below should be taken with caution.

In the 20 day hindcasts, we find no useful correlations between MJO fidelity and either vertical (Figure 3a) or horizontal (Figure 3c) NGMS, or between fidelity and total NGMS (Figure 3e). Models with equal fidelity (e.g., MetUM-GA3, GISS-E2, ECEarth3, and MRI-AGCM3) produce NGMS values of opposite signs. The models with the highest (CAM5-ZM) and lowest (CanCM4) MJO fidelity produce similar horizontal and vertical NGMS values. There is no correspondence between vertical NGMS for an individual GCM in the 20 day hindcasts and 20 year climate simulations (Figure 3b), or for total NGMS (Figure 3f). There is no evidence that, for one GCM, variations in NGMS between the two experiments can account for variations in MJO fidelity. Curiously, there is a negative correlation between horizontal NGMS in the two experiments: GCMs that produce negative horizontal NGMS in 20 day hindcasts (CNRM-AM, CAM5-ZM) show strongly positive horizontal NGMS values in the 20 year climate simulations (Figure 3d). Further investigation of this behavior is outside the scope of this study and may be limited to this set of GCMs or caused by the small sample of 20 day hindcast data.

3.4. Time Step Precipitation Variability

In analyzing the 2 day hindcasts, *Xavier et al.* [2015] discovered substantial time step to time step intermittency in precipitation in some GCMs, even when grid point precipitation was averaged across a $5° \times 5°$ region (75°–80°E, 0°–5°N). Time step intermittency in convection, and hence in heating and moistening increments, could influence the GCM dynamics and interfere with the propagation of atmospheric waves, including the MJO. In a more detailed analysis of MetUM-GA3, *Xavier et al.* [2015] showed that the substantial time step intermittency in convection was not associated with variability in the dynamics but that this resulted in poor dynamics-convection coupling on the shortest temporal scales. By contrast, MIROC5 had little time step intermittency in either convection or dynamical fields (e.g., vertical velocity). This analysis did not address whether time step variability in precipitation affected MJO fidelity.

Here we measure time step intermittency in precipitation as the lag-1 root-mean-square difference (RMSD) in time step precipitation from hours 13–48 of the 2 day hindcasts. Like *Xavier et al.* [2015], we remove the first twelve hours of each hindcast to limit model spin-up; removing the first twenty-four hours did not change the conclusions presented below. We compute the lag-1 RMSD from each 2 day hindcast by (a) area-averaging time step precipitation in 75°–80°E, 0°–5°N; (b) creating a second time series by lagging the time series from (a) by one time step; (c) computing the RMSD between the time series from (a) and from (b). We average the RMSD values across all forty-four 2 day hindcasts.

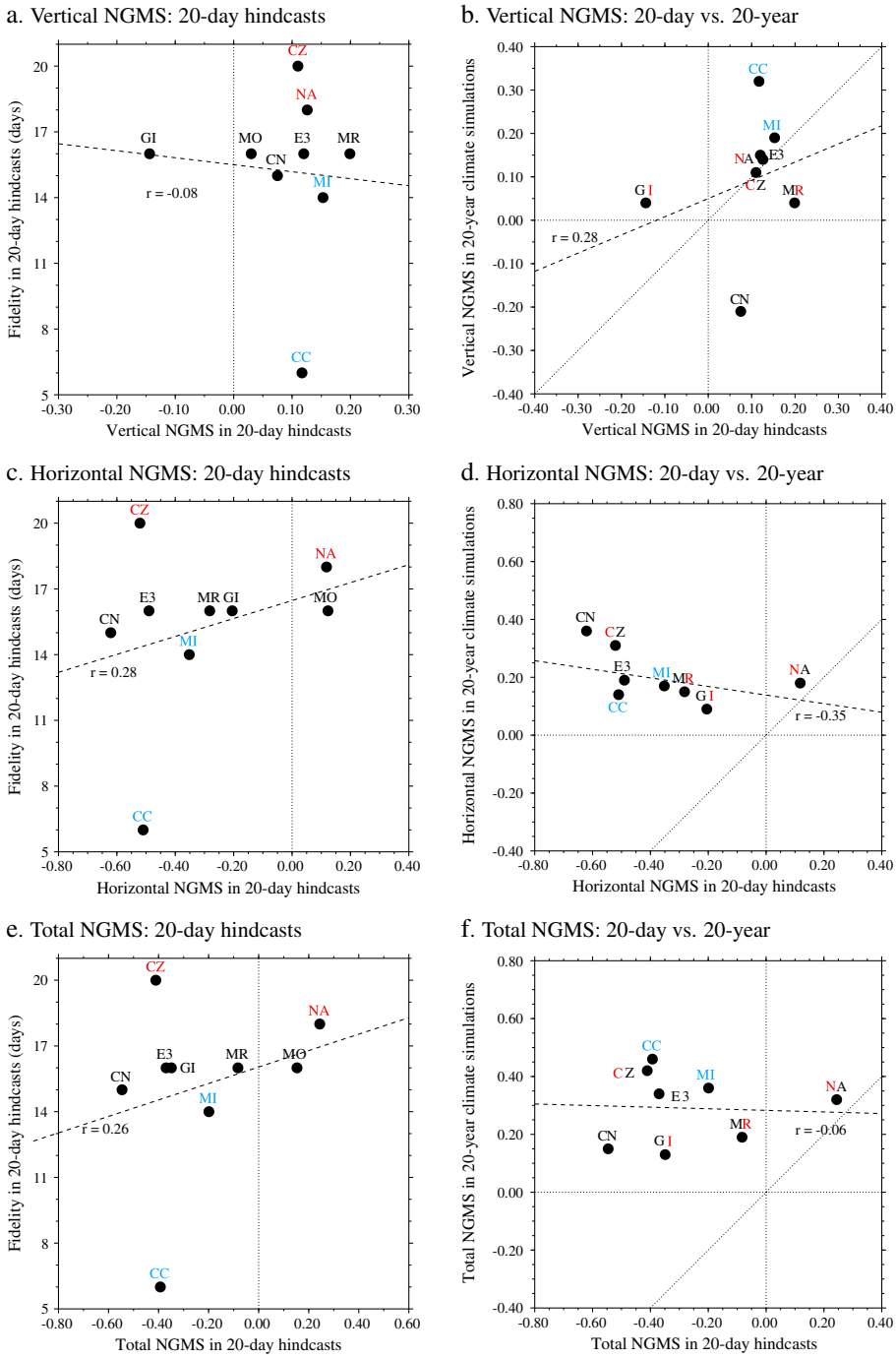

Figure 3. (a,b) Vertical normalized gross moist stability (NGMS), (c,d) horizontal NGMS, and (e,f) total NGMS: (Figures 3a, 3c, and 3e) MJO fidelity in 20 day hindcasts against NGMS; (Figures 3b, 3d, and 3f) NGMS in 20 day hindcasts against NGMS in 20 year climate simulations. The least-squares regression lines and correlation coefficients are also shown; Figures 3b, 3d, and 3f also include a one-to-one line. In Figures 3a, 3c, and 3e, model codes are colored by fidelity in 20 day hindcasts, as in Figure 2a; in Figures 3b, 3d, and 3f, model codes are colored as in Figure 1a. For the 20 day hindcasts, NGMS is computed from only days 3–20.

We compare each model's RMSD values to MJO fidelity in the 20 day hindcasts (Figure 4a) and 20 year climate simulations (Figure 4b). While there is a modest negative correlation between the RMSD values and 20 day hindcast fidelity ($r = -0.43$, $p > 0.20$), this becomes insubstantial if CanCM4 is removed ($r = -0.24$, $p > 0.20$). Likewise, there is no relationship between the RMSD values and MJO fidelity in the 20 year climate simulations

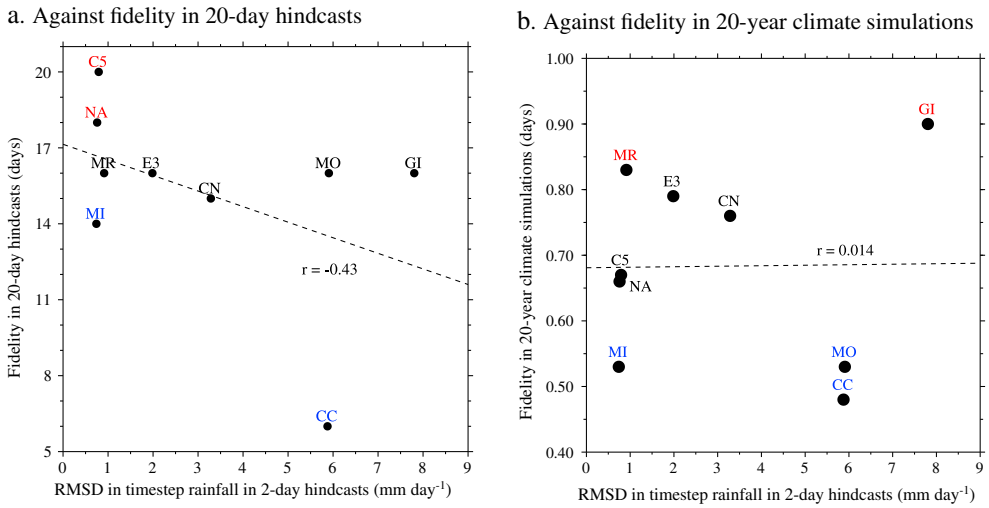

Figure 4. For each model, the lag-1 root-mean-square difference (RMSD) in area-averaged (75°–80°E, 0°–5°N) time step precipitation in the 2 day hindcasts (mm day^{-1}) against MJO fidelity in the (a) 20 day hindcasts and (b) 20 year climate simulations. The least-squares regression lines and correlation coefficients are also shown. Without CanCM4 (CC) the correlation coefficient in Figure 4a is −0.24. Model codes are colored by fidelity in Figure 4a (20-day hindcasts) and Figure 4b (20-year climate simulations), relative to all GCMs for which 20 day hindcasts and 20 year climate simulations, respectively, were submitted.

(r = 0.01, p > 0.20). Four models produce low RMSD values—MRI-AGCM3, CAM5, GEOS5, and MIROC5—but vary widely in their fidelity. Likewise, MRI-AGCM3 and GISS-E2 show the highest fidelity, but MRI-AGCM3 generates smooth time step precipitation, while the precipitation in GISS-E2 is highly intermittent. Among these GCMs, there appears to be no relationship between time step variability in convection and MJO performance. The substantial intermodel variation in time step precipitation intermittency represents an interesting avenue for further research, focusing on the effects on longer temporal scales and interactions with the resolved dynamics.

3.5. Tropospheric Moistening

Klingaman et al. [2015] found that compositing vertical profiles of net moistening by rain rate produced a diagnostic that was most able to distinguish between higher-skill and lower skill GCMs in the 20 day hindcasts, although the distinction was far from absolute. When applied to net moistening and rainfall from ECMWF-YOTC 24 h forecasts (Figure 5a), this diagnostic shows that as precipitation increases, the profile of net moistening transitions from low-level moistening and upper level drying at low rain rates (<2 mm day^{-1}), through to midlevel moistening at moderate rain rates (2–9 mm day^{-1}) to upper level moistening and low-level drying at heavy rain rates (>9 mm day^{-1}), with additional low-level and midlevel drying at the strongest rain rates (>30 mm day^{-1}). Pattern correlations of this diagnostic between GCMs and ECMWF-YOTC produced a significant relationship with hindcast skill for all 20 day hindcast GCMs ($r = 0.82, p \sim 0.01$) [*Klingaman et al.*, 2015]. The authors hypothesized that the midlevel moistening at moderate rain rates was critical to a reliable representation of the MJO. In ECMWF-YOTC and the high-skill GCMs, midlevel moistening was produced by a combination of the GCM dynamics and physics, while in lower skill GCMs moistening tendencies from the dynamics and physics were nearly always of opposite signs (although not of equal magnitudes).

We compute the composite vertical profiles of net moistening by rain rate (hereafter "the net moistening diagnostic") using the method in *Klingaman et al.* [2015]: by compositing net moistening (dq/dt), as well as moistening from the GCM dynamics and physics (summing all physics tendencies) within rain rate ranges over all grid points in a Warm Pool domain: 60–160°E and 10°S–10°N, using data from each experiment at the finest temporal and spatial resolutions captured (section 2.1). From the 2 day hindcasts, we use only hours 13–48 of each hindcast and convert the time step data on the GCM native vertical grid to pressure coordinates using supplied pressure data; we use all days of each 20 day hindcast. We aim to understand whether the correlations between MJO fidelity and the net moistening diagnostic hold for the 20 year climate simulations, as well as to assess whether the moistening-rainfall relationships apply to GCM time step and grid point scales.

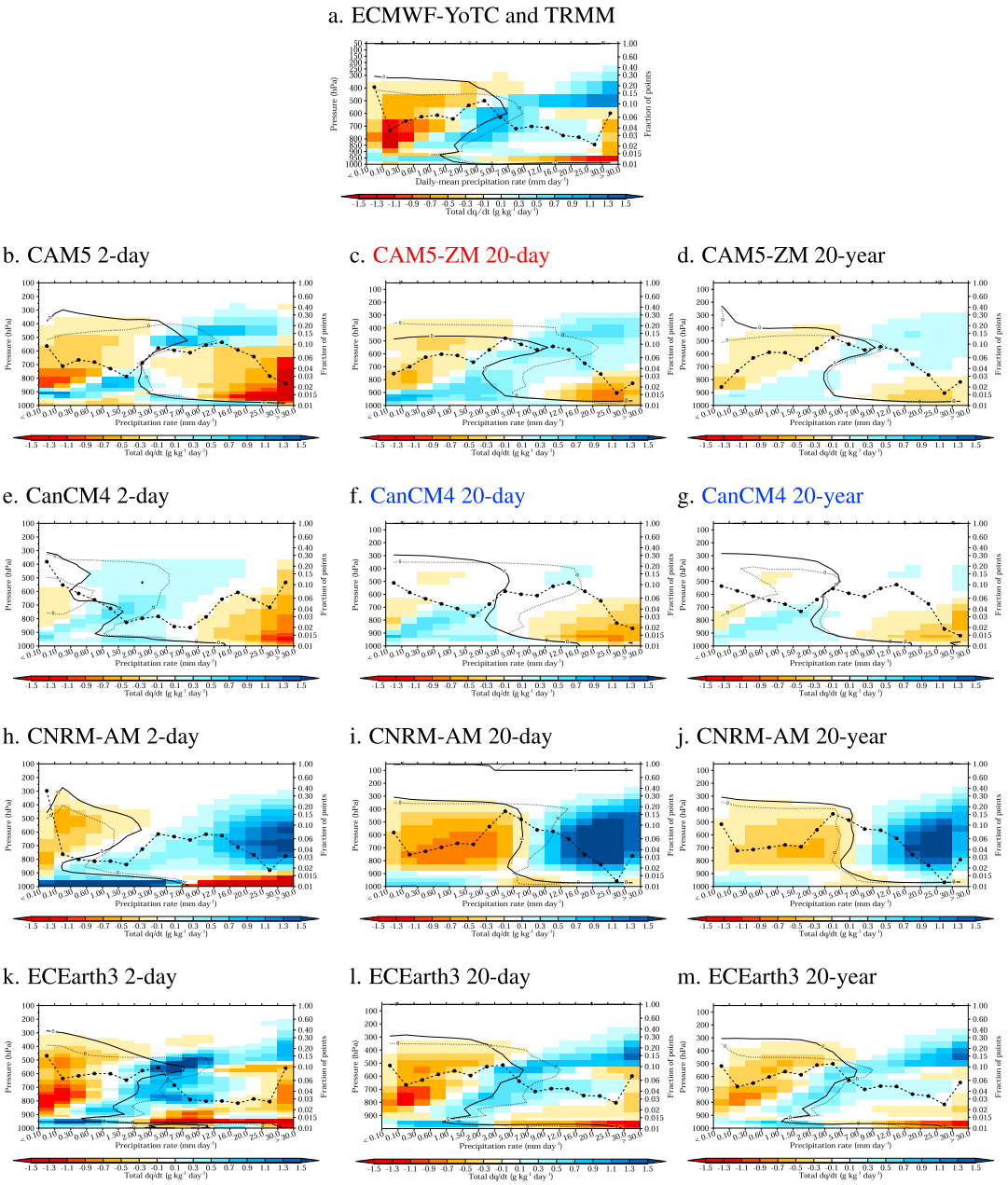

Figure 5. (a–ab) Models and experiments describing data as follows: Shading shows the mean vertical profiles of the rate of change of specific humidity (dq/dt; g kg d^{-1}) for each range of rain rates on the horizontal axis. Solid (dotted) lines are zero contours of the dq/dt from GCM dynamics (physics). Dynamics tendencies are positive (moistening) above and to the right of the solid line; physics tendencies are positive below and to the left of the dotted line. Dashed lines show probability distribution functions (PDFs) of rain rates, using the right-hand vertical axis. Composites are computed from time step data for 2 day hindcasts, 3 h data for 20 day hindcasts, and 6 h data for 20 year climate simulations. For 20 day hindcasts and 20 year climate simulations, colored panel labels indicate MJO fidelity. Note that Figure 5a is repeated for ease of comparison.

While the net moistening diagnostic produces consistent results in the 20 day hindcasts (Figure 5, middle column) and 20 year climate simulations (Figure 5, right column) for most GCMs, there are often substantial discrepancies between the 2 day hindcasts (Figure 5, left column) and the other two experiments. For example, in the 2 day hindcasts, CNRM-AM produces a relatively smooth transition between low-level and upper level moistening (Figure 5h), but this transition becomes very sharp in the 20 day hindcasts (Figure 5i) and 20 year climate simulations (Figure 5j). There is also a marked change in the rain rate PDF, with many more near-zero values in the time step PDF from the 2 day hindcasts than in the 3 h PDF from the 20 day

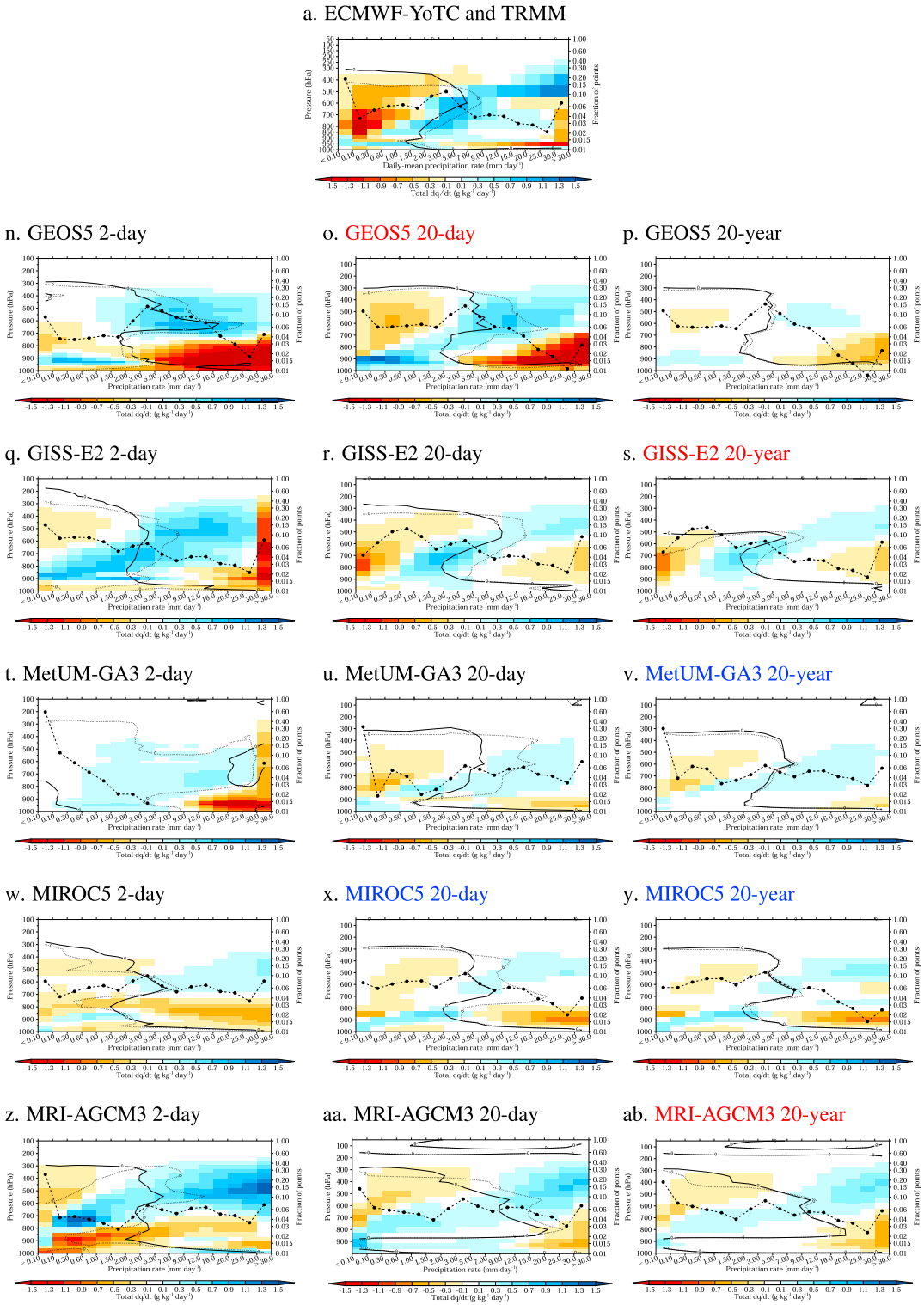

Figure 5. (continued).

hindcasts. This suggests that the rain rate PDFs in the 20 day hindcasts and 20 year climate simulations arise from temporal averaging: the PDF peak at 3–5 mm day^{-1} in the 20 day hindcasts is likely due to averaging several time steps of near-zero precipitation together with one time step of heavy precipitation. As expected, the GCMs that show the largest changes in rain rate PDFs between the time step and temporally averaged

data—CanCM4, CNRM-AM, GISS-E2, and MetUM—also show the strongest time step to time step variability in precipitation (Figure 4). Conversely, GCMs with low time step variability—CAM5, ECEarth3, MIROC5, MRI-AGCM3, and GEOS5—produce consistent precipitation PDFs at the time step, 3 h and 6 h scales.

However, linear combinations of time steps of high and low rainfall cannot explain many of the variations in the composite net moistening profiles between the 2 day hindcasts and the 20 day hindcasts. Again using CNRM-AM as an example, the strong net drying at 750–600 hPa and 2.0–3.0 mm day^{-1} in the 20 day hindcasts (Figure 5i) does not appear in the time step data (Figure 5h) at that height in any rain rate band. Similarly, in the 2 day hindcasts CAM5 does not moisten at 700 hPa for any rain rate (Figure 5b), but CAM5-ZM produces moistening at that level in the 20 day hindcasts (Figure 5c) and 20 year climate simulations (Figure 5d). In the 2 day hindcasts, CanCM4 (Figure 5e), GISS-E2 (Figure 5q) and MetUM-GA3 (Figure 5t) display high frequencies of rain rates >30 mm day^{-1} associated with very strong column drying, which suggests that the models are still adjusting their moisture fields away from the ECMWF-YOTC analysis. This calls into question the validity of the net moistening diagnostic for the 2 day hindcast data, as well as whether the GCM parameterization behavior seen in the 2 day hindcasts is affected by strong model adjustment away from the ECMWF-YOTC analysis. We obtained similar results for the 2 day hindcasts when we used only hours 25–48. Further, we note that ECEarth3, which is based on the ECMWF IFS, shows little change in either the rain rate PDF or the net moistening diagnostic between the 2 day hindcasts and the other two experiments (Figures 5k–5m), supporting the hypothesis that many of the changes in the net moistening diagnostic in other GCMs result from spin-up from a "foreign" analysis.

In every GCM, the amplitude of the net moistening diagnostic decreases from the 20 day hindcasts to the 20 year climate simulations. Some of this decrease may be due to temporal averaging, since the 20 day hindcast (20 year climate simulation) diagnostic is computed from 3 h (6 h) data, yet averaging the 20 day hindcast data to 6 h values produced little change in this diagnostic (not shown). The decrease could result from GCM spin-up in the 20 day hindcasts, but *Klingaman et al.* [2015] noted that there was little variation in this diagnostic with lead time. An alternative hypothesis is that most of the 20 day hindcast start dates include a strong MJO in the initial conditions, while most of these GCMs produce a weaker-than-observed MJO in their 20 year climate simulations. Therefore, the amplitude of the moistening diagnostic may be linked to the amplitude of the MJO, or subseasonal tropical convective variability generally, in the simulation. Indeed, CAM5-ZM (Figures 5c and 5d) and GEOS5 (Figures 5o and 5p) show the largest amplitude reductions and are also the two models that show the largest reductions in MJO fidelity between the 20 day hindcasts and 20 year climate simulations (Figure 1).

As in *Klingaman et al.* [2015], we compute pattern correlations between the net moistening diagnostic for ECMWF-YOTC (Figure 5a) and each GCM, to investigate the relationship between fidelity in this diagnostic and MJO fidelity. We refer to these pattern correlations as the "net moistening metric." The 2 day hindcasts show a weak and insignificant correlation ($r = 0.43$, $p > 0.20$) between the net moistening metric and the pattern correlation of the 2 day lead-time rainfall Hövmoller diagram with TRMM (Figure 6a). This is not surprising, given the intermodel similarity in the rainfall Hövmoller pattern correlations (Table 2) and the intermodel variability in the relationship between net moistening and rainfall (Figure 5). This result suggests that the net moistening metric is not valid so early in the hindcasts, when the model precipitation and moisture fields may still be adjusting to the analysis. There is a strong relationship ($r = 0.80$, $p \sim 0.01$) between the net moistening metric and MJO fidelity in the 20 day hindcasts (Figure 6b), confirming the results of *Klingaman et al.* [2015]. We find a slightly weaker but still significant correlation ($r = 0.69$, $p \sim 0.05$) in the 20 year climate simulations (Figure 6c). Critically, no low-fidelity model scores highly in the net moistening metric for either experiment, nor do any of the higher-fidelity models score poorly, although CNRM-AM performs abnormally poorly in the net moistening metric due to its very sharp transition from drying to moistening throughout the free troposphere around 7 mm day^{-1} (Figure 5i,j).

Figure 6d demonstrates that GCMs that increase in relative MJO fidelity (against all models in the experiment, denoted by the colored model codes) from the 20 day hindcasts to the 20 year climate simulations (MRI-AGCM3 and GISS-E2) also show an increase in the net moistening metric, while GCMs that decrease in relative MJO fidelity (GEOS5, CAM5-ZM, and MetUM3) display a decrease in the net moistening metric. GCMs in which relative fidelity does not change (CanCM4, CNRM-AM, ECEarth3, and MIROC5) show small changes in the net moistening metric, with the possible exception of CanCM4, which produced a very low value in the 20 day hindcasts. This agrees with our earlier qualitative assessment that the GCMs with the largest amplitude

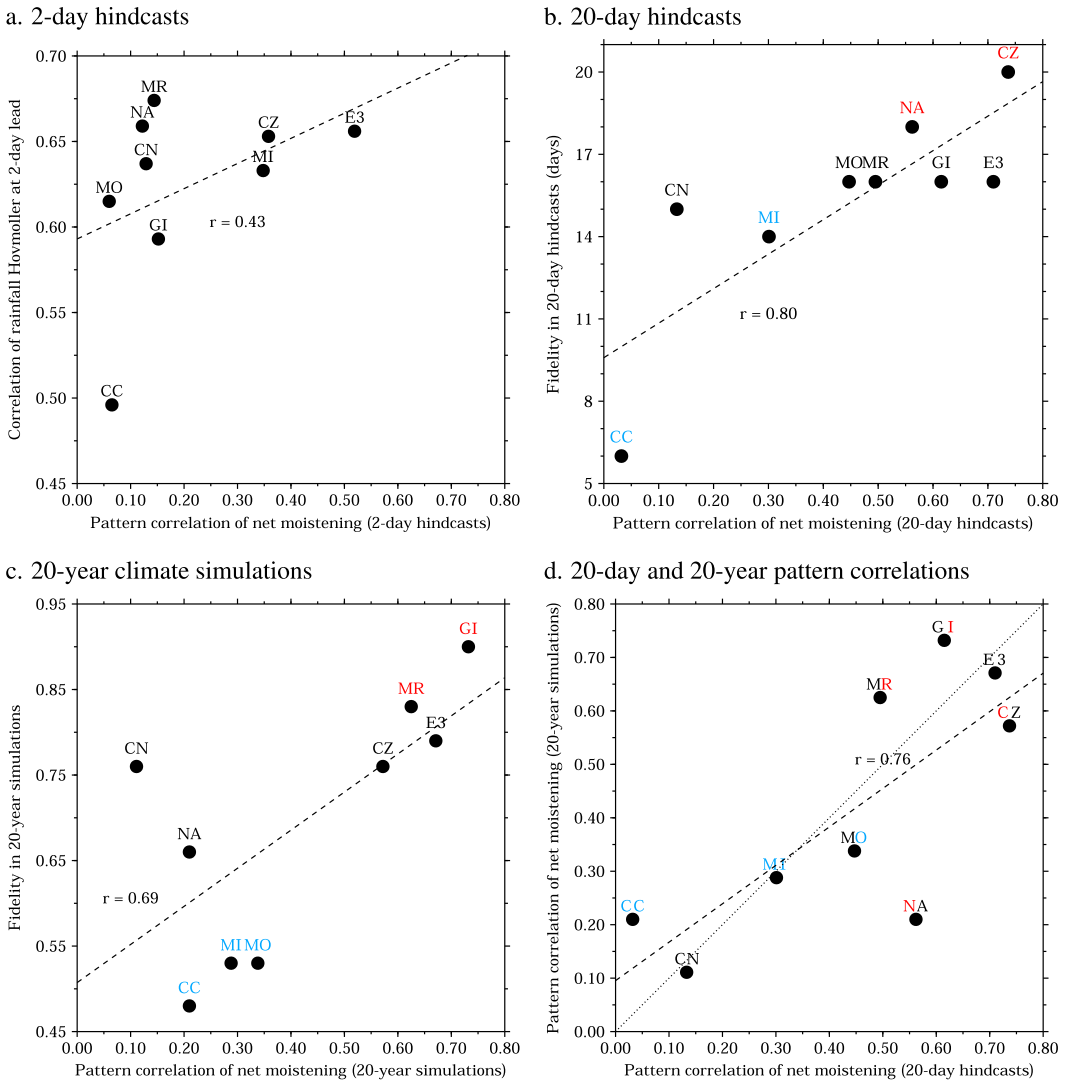

Figure 6. For each GCM, pattern correlations of the net moistening diagnostics are computed against ECMWF-YOTC (Figure 5a). (a) The pattern correlation of net moistening from the 2 day hindcasts against the pattern correlation of the 2 day lead-time rainfall Hövmollers against TRMM (as in Figure 1b); (b) the pattern correlation of net moistening from the 2 day hindcasts against MJO fidelity in the 20 day hindcasts; (c) the same as Figure 6b but for the 20 year climate simulations; (d) comparison of the pattern correlations of net moistening from the 20 day hindcasts and 20 year climate simulations. Colored codes show relative fidelity in (Figure 6b) 20 day hindcasts, (Figure 6c) 20 year climate simulations, and (Figure 6d) 20 day hindcasts for the first letter and 20 year climate simulations for the second letter.

decreases in the net moistening diagnostic between the two experiments also displayed large decreases in MJO fidelity. For this set of GCMs, the net moistening metric distinguishes well between GCMs with relatively higher and lower MJO fidelity; it also explains variations in MJO fidelity between the two experiments.

4. Discussion

The RH difference and net moistening metrics emerge as the measures most able to discern between high- and low-fidelity GCMs across initialized hindcasts and climate simulations, although the relationships with MJO fidelity are far from perfect. These metrics highlight the relationship between precipitation and either total free-tropospheric moisture or its time rate of change. This fits with the recent view that the MJO is a moisture-driven mode of the tropical atmosphere, for which either or both the horizontal and vertical advection of moist static energy by the convectively driven circulation is critical to the maintenance and propagation of the mode [e.g., *Sobel and Maloney*, 2013; *Sobel et al.*, 2014; *Hsu et al.*, 2014]. The RH difference metric emphasizes the sensitivity of simulated rainfall to free-tropospheric moisture, as well as the dynamic

range of RH in the GCM. GCMs that produce strong convection and heavy precipitation in atmospheric columns that are far from saturation will score poorly; in these GCMs, the suppressed MJO phase often resembles a weakened active phase.

The net moistening metric focuses on the height and sign of moisture tendencies for a given rain rate. GCMs that score well produce free-tropospheric drying from subsidence when rain rates are low, low-level and midlevel moistening from advection and convective detrainment as rain rates increase, and upper level moistening from advection and middle- and low-level drying from precipitation at heavy rain rates. As for the RH difference metric, the net moistening metric rewards GCMs that can build and maintain free-tropospheric moisture anomalies, instead of quickly removing them. Although it has not been conclusively demonstrated here, it is likely that low-fidelity GCMs remove these moisture anomalies and limit their dynamic range of RH through convection, which in dry columns detrains moisture into the free troposphere and in moist columns removes it by precipitation. This may be inferred from the zero-moistening contours for the GCM dynamics and physics in Figure 5: in the low troposphere and midtroposphere, the physics tendency is typically of the opposite sign to the net tendency. The exception is in the midtroposphere at moderate rain rates, where in the high-fidelity GCMs the physics contributes to the net moistening, rather than working against the dynamics; *Klingaman et al.* [2015] argued that this was the critical component of the net moistening diagnostic. These results agree with the growing list of studies to show that delaying the response of convection to moisture anomalies improves the MJO in GCMs [e.g., *Wang and Schlesinger*, 1999; *Hannah and Maloney*, 2011; *Hirons et al.*, 2013; *Benedict and Maloney*, 2013; *Klingaman and Woolnough*, 2014], as well as with studies that have concluded that the transition from shallow to midlevel convection is critical to the representation of the MJO [e.g., *Inness et al.*, 2001; *Benedict and Randall*, 2009; *Woolnough et al.*, 2010; *Cai et al.*, 2013]. Finally, we note that the RH difference and net moistening metrics focused, independently, on the low troposphere to midtroposphere (850–500 hPa) as the key region for obtaining the correct relationship between rainfall and moisture.

The 2 day hindcasts were designed to explore GCM parameterization behavior when the models were strongly constrained by a realistic initial analysis. This study has revealed that many diagnostics differ significantly from the 2 day hindcasts to the 20 day hindcasts and 20 year climate simulations. For example, there is little correspondence between GCM skill in precipitation at 2 day and 10 day lead times, as measured by the pattern correlation of rainfall Hövmoller diagrams with TRMM (Figure 1). Further, using only the first two days of the 20 day hindcasts produced unrealistically large NGMS values, as well as RH difference metric values that differed substantially from values obtained from days 3–20. The net moistening diagnostic and rain rate PDFs provided further evidence of different behavior in the 2 day hindcasts, even when only hours 25–48 were examined. While such differences are to be expected, and were the reason for focusing on time step behavior soon after initialization, it does lead one to wonder how these diagnostics evolve as the models drift; it also highlights the challenges in linking the 2 day hindcasts to the 20 day hindcasts and 20 year climate simulations. The large differences in some diagnostics may also suggest that we need to investigate further whether, at 48 h after initialization, the models are still suffering the "shock" of an alien analysis. To help understand the links between the 2 day hindcasts and longer periods, future projects of this nature should also obtain time step data from a selected period during the medium-range hindcasts (at longer leads, e.g., days 9 and 10) and a short period of the climate simulations. Subsequent projects may also consider performing a parallel set of hindcasts in which models are initialized from their own analysis, although this would restrict participation to modeling centers that have an assimilation system.

For many of the nine models considered here, MJO fidelity varies considerably among the 2 day hindcasts, 20 day hindcasts, and 20 year climate simulations (Figure 1). One hypothesis for these variations is that there is no single MJO fidelity measure that can be applied to all three experiments, which makes it impossible to cleanly compare fidelity between the temporal scales and simulation types included here. The hindcast simulations are too short to permit the subseasonal filtering that is applied so often to isolate the MJO in fidelity metrics for climate simulations. While the *Wheeler and Hendon* [2004] RMM indices are often used to evaluate MJO fidelity in initialized hindcasts, there is no accepted RMM-based MJO metric for climate simulations. Many metrics applied to climate simulations identify the MJO with respect to the model's own mean climate, whereas the metrics applied to hindcasts assess how well the model reproduces the observed MJO. We might have classified the models differently if we had computed MJO fidelity in the 20 day hindcasts by projecting each model onto its own leading subseasonal empirical orthogonal functions (EOFs), defined from the 20 year climate simulations, rather than the observed EOFs from *Wheeler and Hendon* [2004].

However, several of the models that performed 20 day hindcasts did not perform 20 year climate simulations. Further, one of the objectives of the 20 day hindcasts was to identify how well the models predicted the observed MJO, not the MJO from the model's own mean climate, so that degradations in prediction skill with lead time could be connected to the growth of biases in simulated physical processes.

A second hypothesis for the variations in MJO fidelity between the experiments is that hindcasts of two MJO events do not adequately sample model behavior in predicting the MJO, which leads to a biased quantification of prediction skill. There is a strong MJO in all (most) of the initial conditions for the 2 day (20 day) hindcasts. A model that is capable of propagating the initial MJO at roughly the observed phase speed, while limiting drift from the analysis to its own mean climate, will perform well in the 2 day and 20 day hindcasts. Yet to perform well in the 20 year climate simulations, a model must also be able to generate an MJO from quiescent conditions. Assessing skill in MJO genesis for these cases would have required a much broader set of hindcast start dates, which would have greatly increased the already-large data burden of the 20 day hindcast experiments. We believe that a fruitful line of further research exists in the connection between a model's ability to generate an MJO "from scratch" in initialized hindcasts and the same model's MJO fidelity in a long climate simulation.

We found that the 20 day hindcasts were an inconvenient length. They were too short to allow some GCMs to drift fully to their intrinsic climatologies, evidenced by the differences in MJO fidelity between the 20 day hindcasts and the 20 year climate simulations (Figure 1), as well as by comparisons of the climatology of the 20 day hindcasts to the mean seasonal cycle of the 20 year climate simulations at the same time of year (not shown). Yet the hindcasts were more than long enough to distinguish the higher-fidelity and lower fidelity GCMs, as shown by RMM bivariate correlations against observations as a function of lead time [see *Klingaman et al.*, 2015, Figure 2a]. The volume of data requested constrained the sample of MJO events (two) and the number of start dates per case (47). The connection between GCM performance in hindcast and climate simulations would have been improved if we had either (a) more start dates per case, particularly if those start dates did not contain an active MJO, so that we could test the ability of the GCMs to generate an MJO when one was not present in the initial conditions; or (b) hindcasts for more MJO cases, so that we could increase our confidence that GCM performance in the hindcasts was representative of the overall GCM skill. If future projects wish to fully examine model drift, they should use hindcasts of longer than 20 days; 30–35 days would likely be sufficient. If such projects aim to more strongly link GCM fidelity in initialized hindcasts and climate simulations, we suggest that they use shorter hindcasts (e.g., 12 days), but a wider range of cases and start dates, including many where the initial RMM amplitude is close to zero, as well as multiple ensemble members.

Even with the substantial quantity of data collected in YOTC, it is clear that we lack the high-frequency, high-resolution, spatially comprehensive observations necessary to validate many of the GCM processes analyzed in this project. This is particularly true at the GCM time step and grid point level, as in the 2 day hindcasts, but also applies to the RH difference and net moistening metrics. While recent campaigns such as the Dynamics of the MJO (DYNAMO) [*Yoneyama et al.*, 2013] have collected high-quality observations of diabatic heating and moistening, the short record of these measurements limits their applicability for GCM development, much in the same way as the limited sample of 20 day hindcasts in this project has limited our ability to connect their behavior to the 20 year simulations. It is difficult to interpret process-oriented diagnostics when applied to short data sets, whether those data sets come from models or observations. For the net moistening metric, we compared the GCMs to ECMWF-YOTC 24 h forecasts because of a lack of comprehensive observations. Those moistening tendencies are fundamentally a model-derived product, even though that model is strongly constrained by its own high-quality analysis. We are fortunate that the tendencies come from a model with a realistic MJO, both in this project [*Klingaman et al.*, 2015] and since 2008 generally [e.g., *Vitart*, 2014]. *Jiang et al.* [2015] compared the RH difference metric in GCMs to TRMM precipitation and ECMWF Interim Reanalysis (ERA-Interim) RH, but TRMM has known issues with detecting light rainfall [e.g., *Huffman et al.*, 2007; *Chen et al.*, 2013] and reanalysis RH is not a substitute for observations. Meaningful progress in improving GCM parameterizations of the diabatic heating, moistening, and momentum mixing by tropical convection requires obtaining high-quality, high-resolution, comprehensive observations of these processes. Specifically, the conclusions of this model evaluation project advocate for high-frequency (i.e., a frequency of an hour or less) observations of precipitation and moistening profiles. These observations must be spatially comprehensive—preferably globally sampled, but at least tropics-wide—and reliable

in both clear-sky and cloudy conditions, to adequately capture the transition MJO phase that our results show may be critical for representing the MJO in models.

Analysis of observations of diabatic moistening from DYNAMO have shown that there is a strong diurnal cycle in lower tropospheric moistening during the suppressed phase of the MJO, with a peak moistening in the afternoon following the peak insolation and the diurnal SST maximum [*Ruppert and Johnson*, 2015]. This moistening is driven by shallow and midlevel convection, but produces little precipitation, presumably because the moisture is detrained quickly into the lower troposphere. It has been hypothesized that this convection may "recharge" tropospheric moisture, priming the atmosphere for the next MJO active phase. Combined with our results that suggest that capturing the relationship between precipitation and moisture is important for the representation of the MJO in models, these observations provide motivation for a model evaluation project focused on the MJO events observed during DYNAMO. We recommend an initialized hindcast experiment for the DYNAMO MJO events, following the suggestions detailed above concerning the length of the hindcasts, the choice of start dates and the collection of time step data at several points in the hindcasts. The DYNAMO hindcast experiment should focus on the diurnal cycle of convection and on validating models against the wealth of observations of diabatic processes collected in DYNAMO. In addition to a set of hindcasts initialized from ECMWF analyses, we recommend that modeling centers that have an analysis system be encouraged to perform hindcasts with the model initialized from its own analysis, to quantify the effects of initializing from the "foreign" ECMWF analysis. In our specification for the 20 day and 2 day hindcasts, we overlooked the specification of the SST boundary condition; although we believe that discrepancies in the SST specification did not affect our conclusions, we recommend that any future experiment specify either persisted initial SSTs or persisted initial SST anomalies on a time-varying climatology.

5. Summary and Conclusions

The "Vertical structure and physical processes of the Madden-Julian oscillation" project has collected and evaluated an extensive data set of output from 32 GCMs, including temperature, moisture, and momentum tendencies from individual subgrid-scale parameterizations. Three experiments were performed that spanned short-range hindcasts, from which time step output was collected, to long climate simulations with subdaily data (section 2.1). The results of those experiments are presented in companion manuscripts [*Klingaman et al.*, 2015; *Jiang et al.*, 2015; *Xavier et al.*, 2015], in which diagnostics and metrics are developed that isolate GCM processes (e.g., diabatic heating and moistening associated with tropical convection) to compare GCMs to one another and to observations, where observations exist. In *Jiang et al.* [2015] and *Klingaman et al.* [2015], these diagnostics were also correlated against measures of MJO fidelity but only for the experiment considered in each study. Since GCM behavior across temporal scales (i.e., between time steps and 6 h averages) and across model background states (i.e., between initialized hindcasts and free-running climate simulations) is a key focus of the project, this manuscript has applied the most discerning diagnostics from each component of the project to data from the set of nine GCMs that contributed to all three experiments.

We find weak and statistically insignificant relationships between MJO fidelity in the 2 day hindcasts, 20 day hindcasts, and 20 year climate simulations (Figure 1). In the 2 day hindcasts, most models show a similar ability to predict the longitude-time pattern of rainfall at a 2 day lead time. The fidelity of these predictions are poorly correlated with fidelity in either the 20 day hindcasts or 20 year climate simulations. Model skill in predicting the *Wheeler and Hendon* [2004] RMM indices in the 20 day hindcasts is also not significantly correlated with MJO fidelity in the 20 year climate simulations. However, we emphasize that these comparisons of MJO fidelity are far from clean, because we lack a single MJO fidelity metric that can be applied identically in all three experiments. Additionally, fidelity in these initialized hindcasts does not require a model to be able to generate an MJO from quiescent conditions, an ability that is necessary to achieve high fidelity in the climate simulations.

In the 20 day hindcasts and 20 year climate simulations, higher-fidelity GCMs tend to score well in the RH difference metric from *Jiang et al.* [2015] (Figure 2) and the net moistening metric from *Klingaman et al.* [2015] (Figure 6), while low-fidelity GCMs tend to score poorly. However, these relationships are far from perfect, as exemplified by the five GCMs that have similar values of the RH difference metric in the 20 day hindcasts, but which vary considerably in MJO fidelity (Figure 2a). In most cases, these metrics can account for differences in relative MJO fidelity (i.e., fidelity relative to all GCMs in that experiment, not only the nine considered here) between the experiments: GCMs that increase (decrease) in relative fidelity from one

experiment to the other also show an increase (decrease) in the metric. These results hold more strongly for the net moistening metric, for which correlations with fidelity are higher, but it is difficult to draw conclusions about the relative worth of these two metrics based on only nine GCMs. These results suggest that to improve the MJO, GCM developers should target the relationship between tropical convection and low-to-middle-tropospheric moisture, including diabatic moistening by convection, rather than the relationship with diabatic heating, which was not associated with MJO fidelity in the 20 day hindcasts [*Klingaman et al.*, 2015]. To test this hypothesis, we recommend a further initialized hindcast experiment, focused on the DYNAMO MJO cases, for which detailed observations of diabatic processes are available.

Despite a negative correlation between both vertical and total NGMS and MJO fidelity in the 20 year climate simulations, we found no correlations between any component of NGMS and MJO fidelity in the 20 day hindcasts (Figure 3). Low correspondence between NGMS in the hindcasts and climate simulations may indicate issues with computing NGMS from the limited sample of hindcast data, however. There was also no correlation between MJO fidelity in the 20 day hindcasts and the time step to time step variability in convection identified by *Xavier et al.* [2015] in the 2 day hindcasts (Figure 4). It was often not possible to apply diagnostics developed for the 20 year climate simulations and 20 day hindcasts to the 2 day hindcasts, due to the short record lengths. When diagnostics were applied, such as the net moistening diagnostic, the results proved difficult to interpret in the context of the other two experiments because of strong drift away from the ECMWF-YOTC analyses. The large changes in parameterization behavior across timescales highlight the challenges of linking the behavior of the GCMs when constrained by an analysis to their behavior as they drift toward their intrinsic climatology.

Finally, we note that the complete data set is available through http://earthsystemcog.org/projects/gass-yotc-mip. While the 2 day hindcast data was archived only over a limited Warm Pool domain, data for the other experiments were collected for at least 50°S – 50°N. We hope that this highly detailed data set will be useful for a variety of tropical and extra-tropical applications beyond analysis of the MJO.

Acknowledgments

All data from all experiments in this project are freely available through http://earthsystemcog.org/projects/gass-yotc-mip. The authors express their gratitude to the nine modeling centers that contributed GCM data to all three experiments in this project, each of which was demanding enough on its own. Comments from Chidong Zhang and an anonymous reviewer helped us to improve the manuscript. Nicholas Klingaman and Steven Woolnough were funded by the National Centre for Atmospheric Science, a collaborative center of the Natural Environment Research Council, under contract R8/H12/83/001. Xianan Jiang acknowledges support by NSF Climate and Large-Scale Dynamics Program under award AGS-1228302 and NOAA MAPP program under award NA12OAR4310075. Duane Waliser acknowledges support from the Office of Naval Research under Project ONRBAA12-001 and NSF AGS-1221013, and the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

Benedict, J. J., and E. D. Maloney (2013), Tropical intraseasonal variability in version 3 of the GFDL atmosphere model, *J. Clim.*, *26*, 426–449.

Benedict, J. J., and D. A. Randall (2009), Structure of the Madden-Julian oscillation in the superparameterized CAM, *J. Clim.*, *66*, 3277–3296.

Benedict, J. J., E. D. Maloney, A. H. Sobel, and D. M. Frierson (2014), Gross moist stability and MJO simulation in three full-physics GCMs, *J. Atmos. Sci.*, *71*, 3327–3349.

Boyle, J. S., S. Klein, G. Zhang, S. Xie, and X. Wei (2008), Climate model forecast experiments for TOGA COARE, *Mon. Weather Rev.*, *136*, 808–832.

Cai, Q., G. J. Zhang, and T. Zhou (2013), Impacts of shallow convection on MJO simulation: A moist static energy and moisture budget analysis, *J. Clim.*, *26*, 2417–2431.

Chen, Y., E. E. Ebert, K. J. E. Walsh, and N. E. Davidson (2013), Evaluation of TRMM 3B42 precipitation estimates of tropical cyclone rainfall using PACRAIN data, *J. Geophys. Res. Atmos.*, *118*, 2184–2196, doi:10.1002/jgrd.50250.

de Boisséson, E., M. A. Balmaseda, F. Vitart, and K. Mogensen (2012), Impact of the sea surface temperature forcing on hindcasts of Madden-Julian oscillation events using the ECMWF model, *Ocean Sci.*, *8*, 1071–1084.

Del Genio, A. D., Y. Chen, Y. Kim, and Y. Mao-Sung (2012), The MJO transition from shallow to deep convection in CloudSat/CALIPSO data and GISS GCM simulations, *J. Clim.*, *25*, 3755–3770.

Hannah, W. M., and E. D. Maloney (2011), The role of moisture-convection feedbacks in simulating the Madden-Julian oscillation, *J. Clim.*, *24*, 2754–2770.

Hazeleger, W., et al. (2012), EC-Earth v2.2: Description and validation of a new seamless Earth system prediction model, *Clim. Dyn.*, *39*, 2611–2629.

Hirons, L. C., P. Inness, F. Vitart, and P. Bechtold (2013), Understanding advances in the simulation of intraseasonal variability in the ECMWF model. Part II: The application of process-based diagnostics, *Q. J. R. Meteorol. Soc.*, *139*, 1427–1444, doi:10.1002/qj.2059.

Hsu, P.-C., T. Li, and H. Murakami (2014), Moisture asymmetry and MJO eastward propagation in an aquaplanet general circulation model, *J. Clim.*, *27*, 8747–8760.

Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff (2007), The TRMM multi-satellite precipitation analysis: Quasi-global, multi-year, combined-sensor precipitation estimates at fine scale, *J. Hydrometeorol.*, *8*, 38–55.

Hung, M.-P., J.-L. Lin, W. Wang, D. Kim, T. Shinoda, and S. J. Weaver (2013), MJO and convectively coupled equatorial waves simulated by CMIP5 climate models, *J. Clim.*, *26*, 6185–6214.

Inness, P. M., J. M. Slingo, E. Guilyardi, and J. Cole (2001), Organization of tropical convection in a GCM with varying vertical resolution: Implications for the simulation of the Madden-Julian Oscillation, *Clim. Dyn.*, *17*, 777–793.

Jiang, X., et al. (2015), Vertical structure and physical processes of the Madden-Julian oscillation: Exploring key model physics in climate simulations, *J. Geophys. Res. Atmos.*, doi:10.1002/2014JD022375.

Kim, D., A. H. Sobel, A. D. Del Genio, Y. Chen, S. J. Camargo, M.-S. Yao, M. Kelley, and L. Nazarenko (2012), The tropical subseasonal variability simulated in the NASA GISS general circulation model, *J. Clim.*, *25*, 4641–4659.

Kim, D., J.-S. Kug, and A. H. Sobel (2014), Propagating vs. non-propagating Madden-Julian oscillation events, *J. Clim.*, *27*, 111–125.

Kim, H.-M., C. D. Hoyos, P. J. Webster, and I.-S. Kang (2008), Sensitivity of MJO simulation and predictability to sea surface temperature variability, *J. Clim.*, *21*, 5304–5317.

Klingaman, N. P., and S. J. Woolnough (2014), Using a case-study approach to improve the Madden-Julian oscillation in the Hadley Centre model, *Q. J. R. Meteorol. Soc.*, *140*, 2491–2505.

Klingaman, N. P., et al. (2015), Vertical structure and physical processes of the Madden-Julian oscillation: Linking hindcast fidelity to simulated diabatic heating and moistening, *J. Geophys. Res. Atmos.*, doi:10.1002/2014JD022374.
Lin, J.-L., K. M. Weickman, G. N. Kiladis, B. E. Mapes, S. D. Schubert, M. J. Suarez, J. T. Bacmeister, and M.-I. Lee (2008), Subseasonal variability associated with Asian summer monsoon simulated by 14 IPCC AR4 coupled GCMs, *J. Clim.*, *21*, 4542–4566.
Madden, R. A., and P. R. Julian (1971), Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific, *J. Atmos. Sci.*, *28*, 702–708.
Madden, R. A., and P. R. Julian (1972), Description of global-scale circulation cells in the tropics with a 40–50 day period, *J. Atmos. Sci.*, *29*, 1109–1123.
Maloney, E. D., X. Jiang, S.-P. Xie, and J. Benedict (2014), Process-oriented diagnosis of East Pacific warm pool intraseasonal variability, *J. Clim.*, *27*, 6305–6324.
Merryfield, W. J., W.-S. Lee, G. J. Boer, V. V. Kharin, J. F. Scinocca, G. M. Flato, R. S. Ajayamohan, J. C. Fyfe, Y. Tang, and S. Polavarapu (2013), The Canadian seasonal to interannual prediction system. Part I: Models and initialization, *Mon. Weather Rev.*, *141*, 2910–2945.
Moncrieff, M. W., D. E. Waliser, M. J. Miller, M. A. Shapiro, G. R. Asrar, and J. Caughey (2012), Multiscale convective organisation and the YOTC virtual global field campaign, *Bull. Amer. Meteor. Soc.*, *93*, 1171–1187.
Morrison, H., and A. Gettelman (2008), A new two-moment bulk stratiform cloud microphysics scheme in the community atmosphere model, version 3 (cam3), *J. Clim.*, *21*, 3642–3659.
Neale, R. B., et al. (2012), Description of the NCAR Atmospheric model: CAM5.0, *Tech. Rep. NCAR/TN-486+STR*, Natl. Cent. for Atmos. Res., Boulder, Colo.
Neelin, J. D., and I. M. Held (1987), Modeling tropical convergence based on the moist static energy budget, *Mon. Weather Rev.*, *115*, 3–12.
Petch, J., D. Waliser, X. Jiang, P. K. Xavier, and S. Woolnough (2011), *A global model intercomparison of the physical processes associated with the Madden-Julian oscillation*. GEWEX News, August, pp. 5.
Pritchard, M. S. (2014), Causal evidence that rotational moisture advection is critical to the superparameterized Madden–Julian oscillation, *J. Atmos. Sci.*, *71*, 800–815.
Raymond, D. J., S. S. Sessions, A. H. Sobel, and Z. Fuchs (2009), The mechanics of gross moist stability, *J. Adv. Model. Earth Syst.*, *1*, 9, doi:10.3894/JAMES.2009.1.9.
Rienecker, M. M., et al. (2008), The GEOS-5 data assimilation system: Documentation of version 5.0.1, 5.1.0 and 5.2.0, *Tech. Rep. Series on Global Modeling and Data Assimilation*, National Aeronautics and Space Administration, Greenbelt, Md.
Ruppert, J. H., and R. H. Johnson (2015), The pre-onset stage of the Madden-Julian oscillation during DYNAMO, *J. Atmos. Sci.*, *72*, 1622–1647, doi:10.1175/JAS-D-14-0218.1.
Schmidt, G., et al. (2014), Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, *J. Adv. Model. Earth Syst.*, *6*, 141–184.
Sobel, A. H., and E. M. Maloney (2013), Moisture modes and the eastward propagation of the MJO, *J. Atmos. Sci.*, *70*, 187–192.
Sobel, A. H., S. Wang, and D. Kim (2014), Moist static energy budget of the MJO during DYNAMO, *J. Atmos. Sci.*, *71*, 4276–4291.
Song, X., and G. J. Zhang (2011), Microphysics parameterization for convective clouds in a global climate model: Description and single-column model tests, *J. Geophys. Res.*, *116*, D02201, doi:10.1029/2010JD014833.
Song, X., G. J. Zhang, and J.-L. F. Li (2012), Evaluation of microphysics parameterization for convective clouds in the NCAR Community Atmosphere Model CAM5, *J. Clim.*, *25*, 8568–8590.
Thayer-Calder, K., and D. A. Randall (2009), The role of convective moistening in the Madden-Julian oscillation, *J. Atmos. Sci.*, *66*, 3297–3312.
Vitart, F. (2014), Evolution of ECMWF sub-seasonal forecast skill scores, *Q. J. R. Meteorol. Soc.*, *683*, 1889–1899.
Voldoire, A., et al. (2013), The CNRM-CM5.1 global climate model: Description and basic evaluation, *Clim. Dyn.*, *40*, 2091–2121.
Waliser, D. E., et al. (2012), The "Year" of tropical convection (May 2008–April 2010), *Bull. Am. Meteorol. Soc.*, *93*, 1189–1218.
Walters, D. N., et al. (2011), The MET office unified model global atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations, *Geosci. Model Dev.*, *4*, 919–941.
Wang, W., and M. E. Schlesinger (1999), The dependence on convective parameterization of the tropical intraseasonal oscillation simulated by the UIUC 11-layer atmospheric GCM, *J. Clim.*, *12*, 1423–1457.
Watanabe, M., et al. (2010), Improved climate simulation by MIROC5: Mean states, variability and climate sensitivity, *J. Clim.*, *23*, 6312–6335.
Wheeler, M. C., and H. H. Hendon (2004), An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction, *Mon. Weather Rev.*, *132*, 1917–1932, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.
Williams, K. D., A. Bodas-Salcedo, M. Déqué, S. Fermepin, B. Medeiros, M. Watanabe, C. Jakob, S. A. Klein, C. A. Senior, and D. L. Williamson (2013), The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate mdoels, *J. Clim.*, *26*, 3258–3274.
Woolnough, S. J., et al. (2010), Modelling convective processes during the suppressed phase of a Madden-Julian oscillation: Comparing single-column models with cloud-resolving models, *Q. J. R. Meteorol. Soc.*, *136*, 333–353.
Xavier, P. K. (2012), Intraseasonal convective moistening in CMIP3 models, *J. Clim.*, *25*, 2569–2577.
Xavier, P. K., et al. (2015), Vertical structure and physical processes of the Madden-Julian oscillation: Biases and uncertainties at short range, *J. Geophys. Res. Atmos.*, doi:10.1002/2014JD022718.
Yoneyama, K., C. Zhang, and C. N. Long (2013), Tracking pulses of the Madden–Julian oscillation, *Bull. Am. Meteorol. Soc.*, *94*, 1871–1891.
Yukimoto, S., et al. (2012), A new global model of meteorological research institute: MRI-CGCM3-model description and basic performance, *J. Meteorol. Soc. Jpn.*, *90A*, 23–64.
Zhang, C. (2005), Madden-Julian oscillation, *Rev. Geophys.*, *43*, RG2003, doi:10.1029/2004RG000158.